From: Thayer, Kris [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP
                    (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=3CE4AE3F107749C6815F243260DF98C3-THAYER, KRI]
Sent: 3/3/2018 12:01:47 PM
To: Cawley, Michelle [Michelle.Cawley@icf.com]
CC: Henning, Cara [Cara.Henning@icf.com]; Hubbard, Heidi [Heidi.Hubbard@icf.com]; Soto, Vicki
                    [/o=ExchangeLabs/ou=Exchange Administrative Group
                    (FYDIBOHF23SPDLT)/cn=Recipients/cn=a960b18ed5c440d69eb9db21b8565f97-Soto, Vicki]
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

I think its fine if Heidi coordinates with them to attend. Thanks!

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Friday, March 2, 2018 11:36 AM
To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>; Hubbard, Heidi <Heidi.Hubbard@icf.com>; Soto, Vicki
<Soto.Vicki@epa.gov>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Hi Kris,

When we first discussed the plan to meet with IRIS team to discuss text analytics we talked about including OPPT in this
meeting. Now that we're scheduled for April 4, please let us know if this still makes sense to you.

I included Heidi Hubbard on this email, who manages our OPPT WA. If you agree that it would be helpful to include their
staff, Heidi can coordinate with them to attend the meeting. Or if you would prefer to initiate the invite -- that works
also! Let us know what you think either way.

Thank you!
Michelle

From: Soto, Vicki [mailto:Soto.Vicki@epa.gov]
Sent: Friday, February 16, 2018 8:02 AM
To: Thayer, Kris <thayer.kris@epa.gov>; Cawley, Michelle <Michelle.Cawley@icf.com>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

HI Michelle.
We have Wed April 4 – would that work for you? Our regular meeting time is 9:30 to 10:30. You are welcome to come
in person – we will also broadcast through webinar for all staff.

Thanks
Vicki

From: Thayer, Kris
Sent: Friday, February 16, 2018 6:27 AM
To: Cawley, Michelle <Michelle.Cawley@icf.com>; Soto, Vicki <Soto.Vicki@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Congratulations! That's great news! Please d share the poster!

Vicki – can we lock in a date for April

Hi Kris,

I just wanted to touch base and let you know that our SOT abstract "Simulating Time Saved for Manual Screening of Literature for IRIS Assessments Using Text Analytics" was selected as one of the Top 10 Best Risk-Assessment Related Abstracts. This was such a nice surprise! I know you weren't able to join as a co-author but we definitely appreciate your support for the idea to use IRIS data and run these simulations. I'll share the poster with you when it's completed and look forward to seeing you in San Antonio.

I know it's unlikely we'll get the text analytics meeting scheduled before SOT, but please let us know if April still seems feasible.

Thank you!
Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Saturday, January 20, 2018 5:20 AM
To: Cawley, Michelle <Michelle.Cawley@icf.com>; Soto, Vicki <Soto.Vicki@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Thanks Michelle – we are currently madly prepping for NAS meeting, so Vicki may be a bit slow to respond. Also, it's possible we may be looking into March or April….

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Friday, January 19, 2018 3:07 PM
To: Thayer, Kris <thayer.kris@epa.gov>; Soto, Vicki <Soto.Vicki@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Thanks so much Kris!

Vicki, we look forward to getting something on your calendar. Let us know when you have some dates in mind.

Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Monday, January 15, 2018 8:43 PM
To: Cawley, Michelle <Michelle.Cawley@icf.com>; Soto, Vicki <Soto.Vicki@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Michelle and Cara – CONGRATS!!

Vicki – can you please work with Michelle and Cara to give a demo on DoCTER? Preferably in February. Once we pick a time we can see if it works for OPPT, but we will probably want to record the session and store it on our automation resource page.

Thanks!

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Friday, January 12, 2018 3:21 PM
To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Hi Kris,

I wanted to share that we have published a DoCTER methods paper that specifically details our ensemble supervised clustering method. Please let us know if have any follow-up on the publication and obviously please feel free to share!

Also, I wanted to follow-up on scheduling a text analytics meeting with IRIS staff. The Wednesday 9:30am time slot seems like a good place to start. I'm not sure what your "on weeks" are, but we could do, February 7, 14, or 28. Shall we pick a week first and then determine if we can get OPPT involved at the same time? Cara and I are planning to come up to DC for the meeting as we have tentative plans for a trip in February so we're hoping one trip will work for multiple purposes. Any chance, you will be visiting DC in February? If so, we'll coordinate our trip with yours if possible.

Thank you!
Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Thursday, December 07, 2017 6:04 AM
To: Cawley, Michelle <Michelle.Cawley@icf.com>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract and Scheduling Text Analytics Meeting

Congrats!

I would say look to schedule something for after February given some January deadlines here. We have a standing all scientist meeting every other Wednesday at 9:30-10:30 – maybe that time slot could work? Let me know what you think and I can check in to see if key OPPT staff have any standing meeting conflicts at that time, then we can hone in on a date.

Thanks so much for the update!

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Wednesday, December 6, 2017 4:30 PM
To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: SOT Abstract and Scheduling Text Analytics Meeting

Hi Kris,

I wanted to circle back to let you know that our SOT abstract on using text analytics for IRIS searches was accepted.

I had a reminder on my calendar to reach out to schedule a meeting with your team in first quarter, but I'm thinking makes sense to schedule that in January given the holidays are nearly upon us. Also, not to mention the threat of a government shutdown that is looming!

Let me know what you think.

Thank you,
Michelle

From: Cawley, Michelle
Sent: Friday, October 06, 2017 8:42 AM
To: 'Thayer, Kris' <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract Ready for EPA Clearance

Hi Kris,

Yes, of course! We completely understand and waiting until OPPT has some breathing room also makes sense. We'll touch base again before the end of the year to see when in the first quarter might be a good time.

Thank you,
Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Friday, October 06, 2017 3:51 AM
To: Cawley, Michelle <Michelle.Cawley@icf.com>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract Ready for EPA Clearance

Honestly, with respect to timeline I'm thinking calendar year 2018. Everyone is completely up to eyeballs. Can that work? We have all scientists meeting every other Wednesday 9:30-10:30....perhaps we can use that slot – but after OPPT gets their documents out – that pressure causes them to have to skip optional meetings

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Thursday, October 5, 2017 9:34 AM
To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract Ready for EPA Clearance

Dear Kris,

Yes, I see your point about clearance -- I forgot that Monday was a federal holiday. I'll remove you from the author list, but certainly happy to have any input from you on the abstract or poster (when it's done). ICF will be satisfied with "Kris clearance". My main objective here was to make sure you're aware of what analysis we're doing with IRIS references. Now that we have simulation code, it's really easy to test the efficacy of using DoCTER compared to manual curation.

To your second point on seminar with IRIS and OPPT -- absolutely! We'd love the opportunity to pass on what we've found to your team and I'm sure OPPT would be interested also. Do you have thoughts on timeframe? Cara and I would likely come to DC for it, so would be great to get it scheduled for a time when you're there also.

Thanks!

Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Thursday, October 05, 2017 6:50 AM
To: Cawley, Michelle <Michelle.Cawley@icf.com>
Cc: Henning, Cara <Cara.Henning@icf.com>
Subject: RE: SOT Abstract Ready for EPA Clearance

Michelle,

Thanks so much for sharing! And thanks for adding me as an author but I'm wondering if I should be removed since it's CRAZY this week and next so I'm concerned about our EPA clearance process (Monday is a holiday) – also, I didn't do any of the work:) I really don't want to slow you down.

Let me know if that would be okay.

Separately, perhaps we can set up a seminar with IRIS Program to talk about the findings? Perhaps include OPPT?

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Wednesday, October 4, 2017 10:34 AM
To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>; Varghese, Arun <Arun.Varghese@icf.com>; Hubbard, Heidi <Heidi.Hubbard@icf.com>
Subject: SOT Abstract Ready for EPA Clearance

Dear Kris,

The attached abstract for SOT is ready for EPA clearance. The submission deadline is Monday, October 9.

The results looked great! We reduced LOE while maintaining high recall in all simulations. For chemicals with very few relevant hazard studies (e.g., t-butanol, n-butanol, RDX) using a two-step approach beginning with a chemical-agnostic seed set derived from substances with many relevant results (e.g., arsenic and others) we also achieved recall above 90% and reduced LOE. We're especially excited about the chemical-agnostic approach and look forward to developing this process further.

Thank you for collaborating on this. We look forward to your comments!

Michelle

From: Thayer, Kris [mailto:thayer.kris@epa.gov]
Sent: Friday, September 29, 2017 1:49 PM
To: Cawley, Michelle <Michelle.Cawley@icf.com>
Cc: Henning, Cara <Cara.Henning@icf.com>; Varghese, Arun <Arun.Varghese@icf.com>; Hubbard, Heidi <Heidi.Hubbard@icf.com>
Subject: RE: Proposal for SOT Abstract

No objections from me! Looking forward to seeing the results!

From: Cawley, Michelle [mailto:Michelle.Cawley@icf.com]
Sent: Thursday, September 28, 2017 4:45 PM

To: Thayer, Kris <thayer.kris@epa.gov>
Cc: Henning, Cara <Cara.Henning@icf.com>; Varghese, Arun <Arun.Varghese@icf.com>; Hubbard, Heidi <Heidi.Hubbard@icf.com>
Subject: Proposal for SOT Abstract

Hi Kris,

Cara suggested I reach out as we have an idea for an SOT poster using public IRIS data that we'd really like to collaborate with you on. Arun wrote some code that allows us to simulate how well we could have identified the final list of hazard ID studies if we had used machine learning. We can answer this question for a range of IRIS chemicals by determining how efficiently we are able to identify what was ultimately tagged to hazard ID from the overall literature search results (or an approximation of a broad literature search where full results are not posted in HERO).

After reviewing the publicly available IRIS pages there are 14 substances (includes 5 phthalates) that could be easily run through the simulation code to answer the following questions:

- What is the work saved (elimination rate of non-relevant documents – also known as specificity) when we target a recall of 95%?
- Are our predictions of recall and specificity biased?
- How does the number of seeds/positive studies affect results?
- Can we improve performance (i.e., maximize recall and minimize number of studies to screen) using a hybrid approach of supervised clustering + machine learning?
- In sparse datasets, where there are few relevant studies (e.g., RDX, t-butanol) can we use a generic agnostic-seed set to identify relevant studies in the unclassified corpus?

Please let us know what you think. Our analytical methods will be transparently presented and the results displayed may be replicated by third-parties, given that the datasets are public. The deadline for submission is October 9th and if you're interested we could get an abstract for you to put through clearance by next Tuesday.

Thank you,
Michelle

MICHELLE CAWLEY | Senior Technical Specialist | +1.919.293.1659 (o) | michelle.cawley@icf.com | icf.com
ICF | 2635 Meridian Parkway, Suite 200, Durham, NC 27713 USA | 919.604.4815 (m)
Connect with us on social media.